

This week:

1. PPT 和 evaluation:

本周做了论文 evaluation 的 slides 并进行了 idea 评估, 收集了一些有价值的意见, 整理如下:

- a) POI 的概率函数可以考虑更多方面, 包括统计, 权重, 稀有程度, 微博的“我在。。。”标签。
- b) 考虑地标性建筑
- c) 建立 POI, 基站, 轨迹的对应关系。
- d) 前端设计中, 外面的边框和其颜色展示效果并不是很明显。

2. 设计稿:

本周进行了新项目前端的设计, 包括查询界面设计, 查询领域知识添加面板, 轨迹设计, 语义分析面板设计, 并撰写文档上传。

本周还和老师们讨论了关于新项目的贡献点, 和明洁学长讨论了项目的前端目标, 解释了系统的功能, 并对后台 spark 的应用进行了学习。

3. 贡献点: 总结了新工作的贡献点和论文 tex:

- a) 利用 POI 数据和路网数据对低采样率的轨迹数据进行语义赋予, 克服了普通用户对于低分辨率低精度数据的查询理解瓶颈。
- b) 利用文本处理方法对符合查询要求的轨迹进行不确定性的查询, 量化, 排序
- c) 通过添加领域知识对这种不确定性的轨迹进行可视分析。
- d) 对轨迹进行多角度的语义展示方法

Tex 文本:

Trajectories of moving objects are captured all the time as the progress of sensor technology. On the one hand, trajectory data is massive and spatio-temporal so that there is an urgent need for an effective way to manage these data.

On the other hand, the visualization of the moving object focuses on the display of trajectories, research on uncertainty trajectory usually shows how large the uncertainty is.

Few research show what lead to such uncertainty.

Our work introduces a method that utilize POI data to assign trajectories with semantic information and build index based on textual information to achieve effectively query operation without consider the topology of the road network.

Our system use text similarity and text analysis method to analyze the trajectory and propose an integrated interface that not only provides fast trajectory queries, but also provides the display of the uncertainty-semantic trajectories and their causes.

把贡献点总结为三:

- a) 一种可以让用户定义领域知识, 利用 POI 数据和路网数据对低采样率的轨迹数据进

行语义赋予，对轨迹进行语义赋予的文本索引方式

- b) 一种可以对文本语义轨迹进行自然语言查询的方法，并对查询结果进行不确定性量化方法
- c) 一个可以提供查询展示探索低分辨率低精度的城市轨迹数据的可视分析系统，克服了普通用户对于低分辨率低精度数据的查询理解瓶颈。

4. 论文：

- **SemanticTraj: A New Approach to Interacting with Massive Taxi Trajectories**

再次阅读了这篇文章，本次阅读主要是为了考虑新文章的写作，写作框架和语句，因为这个文章的思路很清晰，我打算参考本文的框架进行写作，并完成了 introduction 的构思，先提出问题，然后解决方案，贡献点。

- **Searching for Data Sources for the Semantic Enrichment of Trajectories**

本文介绍了一种创新的方法，能够对轨迹数据集进行排名，它依赖于监督的方法来学习被访问地点和有意义的数据集之间的关联规则。利用不同数据集之间的信息进行比较，补全不同数据集的方法，然后进行排名。我们的方法中 POI 的数据虽然不能看作是一种轨迹数据，但是也是时空数据之间的补全。本文中应用的方法实际上还是应用了 POI 的种类的特征进行提取分类等。

- **How Space Structures Language**

本文介绍了空间位置和语言文本的关系，通过准确的，直观的文本的直接指示方向可以很好地理解空间位置的信息，空间位置也往往能够对文本产生更高的相似度。

- **IntelliJ, scala, spark 学习与安装**

阅读了一些教程，完成了 spark 后台环境配置和 scala 的入门教程，并联系学长要到了示例代码，开始进行数据处理操作。